RSAC 2025 Conference

Many Voices.
One Community.

SESSION ID: LAW-T10

Can Courts Survive the Deepfake Deluge? Challenges and Potential Solutions

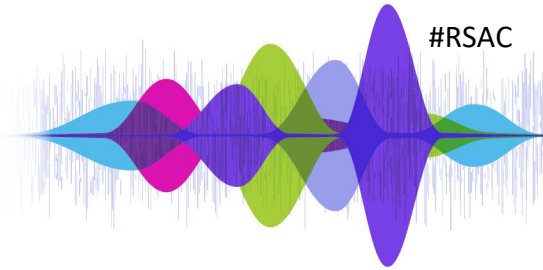
MODERATOR: **Lucy L. Thomson, JD, CISSP**
Principal, Livingston PLLC
ABA Task Force on Law and AI

PANELISTS: **Dr. Maura R. Grossman**
Research Professor, University of Waterloo; Principal, Maura Grossman Law

Hon. Allison H. Goddard
Magistrate Judge
United States District Court
Southern District of California

Hon. Yvonne E. Campos
Superior Court of California
Trial Judge, former Assistant United States Attorney SDCA

Disclaimer



Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

RSAC 2025 Conference

#RSAC

Many Voices.
One Community.

MODERATOR:

Lucy L. Thomson

Livingston PLLC
Principal, Livingston PLLC
ABA Task Force on Law and AI

PANELISTS:

Dr. Maura R. Grossman

Research Professor, University of Waterloo; Principal, Maura Grossman Law

Hon. Allison H. Goddard

Magistrate Judge
United States District Court
Southern District of California

Hon. Yvonne E. Campos

Trial Judge, Superior Court of California, San Diego; former Assistant United States Attorney, SDCA

Agenda

- Deepfake threat landscape
- Impact of deepfakes on the judicial system
- Role of judges, jurors, and experts in evaluating the authenticity of AI-generated evidence
- State of technical solutions that judges and experts might use in court

RSAC 2025 Conference

Looming deepfake threat to the courts as judges grapple with authenticity, validity, reliability, and potential bias of AI-generated evidence

Many Voices.
One Community.

RSAC 2025 Conference

Deepfake threat landscape: national security, judicial system, elections, public trust, disinformation

Contextualizing Deepfake Threats to Organizations

Deepfakes Pose a National Security Threat

Utilizes AI/ML to create believable and highly realistic media.

Most substantial threats from the abuse of synthetic media: techniques that –

- Threaten an organization's brand,
- Impersonate leaders and financial officers, and
- Use fraudulent communications to enable access to an organization's networks, communications, and sensitive information.

FBI: Synthetic Content

The broad spectrum of generated or manipulated digital content, which includes images, video, audio, and text.

These techniques are known as deepfakes or GANs (generative adversarial networks).

https://www.cisa.gov/news-events/alerts/2023/09/12/nsa-fbi-and-cisa-release-cybersecurity-information-sheet-deepfake-threats

Public Service Announcement

FEDERAL BUREAU OF INVESTIGATION

June 5, 2023

Alert Number
I-060523-PSA

Questions regarding this PSA should be directed to your local **FBI Field Office**.

Local Field Office Locations:
www.fbi.gov/contact-us/field-offices

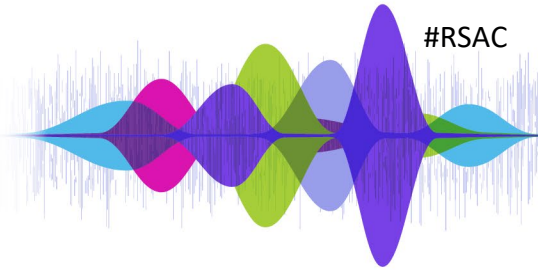
Malicious Actors Manipulating Photos and Videos to Create Explicit Content and Sextortion Schemes

The FBI is warning the public of malicious actors creating synthetic content (commonly referred to as "deepfakes"[a]) by manipulating benign photographs or videos to target victims. Technology advancements are continuously improving the quality, customizability, and accessibility of artificial intelligence (AI)-enabled content creation. The FBI continues to receive reports from victims, including minor children and non-consenting adults, whose photos or videos were altered into explicit content. The photos or videos are then publicly circulated on social media or pornographic websites, for the purpose of harassing victims or sextortion schemes.

RSAC 2025 Conference

AI comes to court

AI comes to court – deepfakes



Acknowledged AI Evidence

1) Evidence is the product of an AI system
2) Expert discloses they used an AI tool in an expert report
3) Video or other exhibits "enhanced by AI"

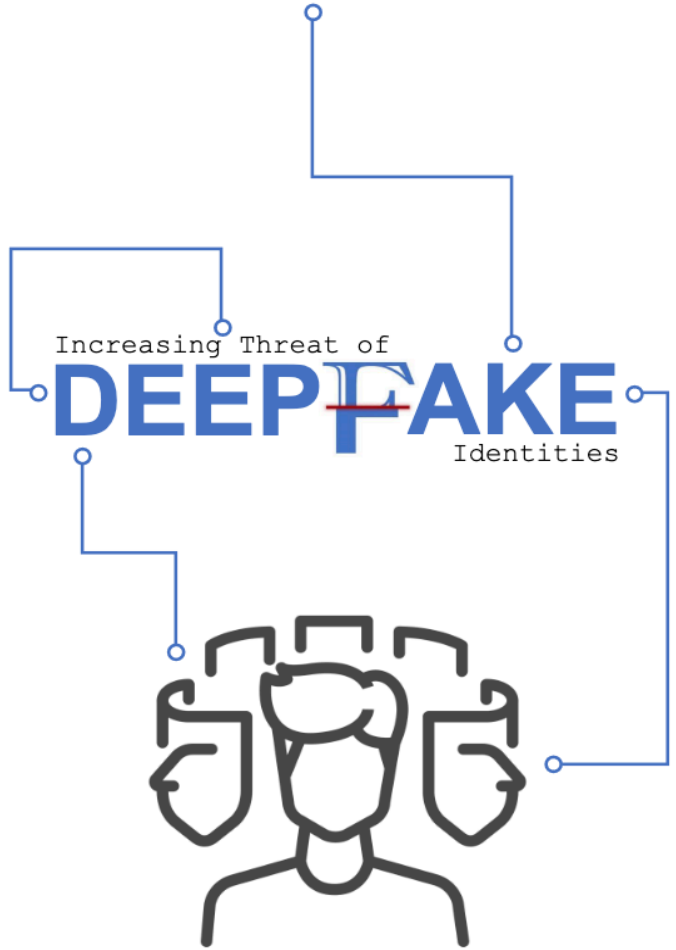
Unacknowledged AI Evidence

(e.g. deepfakes)

4) One party claims the evidence is a real audio/video, other side says it is a deepfake.

National Security & Law Enforcement Scenario 4: Producing False Evidence in a Criminal Case

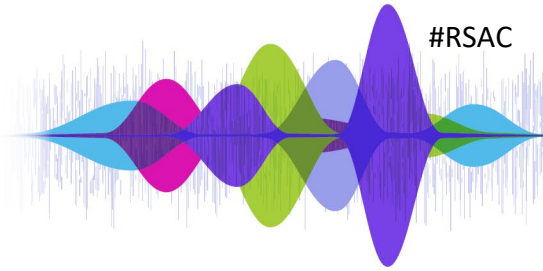
- In this scenario, a wealthy criminal defendant is accused of murder in a building he owns based on evidence including latent fingerprints, hair, DNA and motive.
- Defendant objected to the use of identity verification from videos captured in the building lobby based on low resolution and lack of clarity of the face image.
- As an alibi, submitted to the court video imagery from another location in the building that irrefutably puts him elsewhere at the time the crime took place.
- The defendant's goal is to weaken the biometric evidence and offer contradictory proof that the defendant has a clear alibi. The deepfake content here is the submitted video itself.
- It is not only the timestamp that can be modified to provide an alibi, but unique circumstances can be created in the video so that it looks genuine. The deepfake video, indirectly, also could also render otherwise strong evidence (such as biometrics) circumstantial, due to the specifics of the building and the scene.
  - Page 20



Homeland Security

https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf

Courts focus on AI-generated evidence



- **Authenticity**
- Fed. R. Evid. Rule 901. Authenticating or Identifying Evidence
- (a) In General. To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is.
- **Validity**
- **Reliability**

Prejudicial Effects

Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery

Rebecca A. Delfino

- Taurus Myhand
- Showed subjects altered video; subjects themselves admitted to cheating when they did not!

RSAC 2025 Conference

Impact of deepfakes on the judicial system

RSAC 2025 Conference

Role of judges, jurors, and experts in evaluating the authenticity, validity, and reliability, and potential bias of AI-generated evidence

Many Voices.
One Community.

Grimm and Grossman Proposed Revision to Fed. R. Evid. 901(b)(9) for Acknowledged AI-Generated Evidence

[901](b) **Examples.** The following are examples – not a complete list – of evidence that satisfies the requirement of [901(a)]:

(9) *Evidence about a process or system*. For an item generated by a process or system:

(A) evidence describing it and showing that it produced ~~an accurate~~ a valid and reliable result; and

(B) if the proponent concedes that the item was generated by artificial intelligence, additional evidence that:

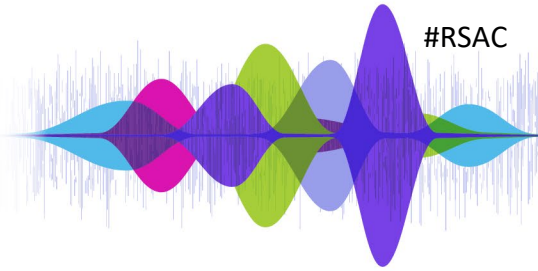
(i) describes the training data and software or program that was used; and

(ii) shows that they produced valid and reliable results in this instance.

Grimm and Grossman Proposed New Fed. R. Evid. 901(c) for Disputed or Potentially Fabricated Evidence

901(c): Potentially Fabricated or Altered Electronic Evidence.

If a party challenging the authenticity of computer-generated or other electronic evidence demonstrates to the court that a jury reasonably could find that the evidence has been altered or fabricated, in whole or in part, using Artificial Intelligence, the evidence is admissible only if the proponent demonstrates that its probative value outweighs its prejudicial effect on the party challenging the evidence.



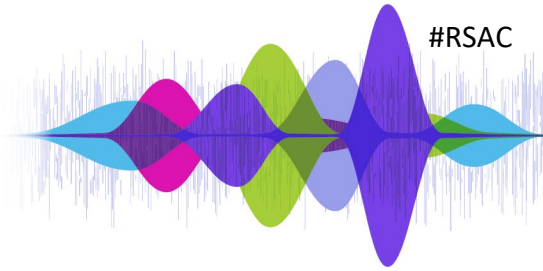
Precedent for Grimm and Grossman's Approach

- ***Huddleston v. US*, 108 S. Ct. 1496, 1502 (1988)**

  The Supreme Court set forth the evidentiary standard judges must apply when determining whether to admit "other crimes, wrongs, or acts" evidence pursuant to Fed. R. Evid. 404(b)

- ***Johnson v. Elk Lake School District*, 283 F.3d 238, 143-33 (3d Cir. 2002)**

  The Third Circuit follows Huddleston when determining whether, in a civil case seeking damages for sexual assault, evidence that the defendant had previously committed an "offense of sexual assault should be admitted pursuant to Fed. R. Evid. 415.



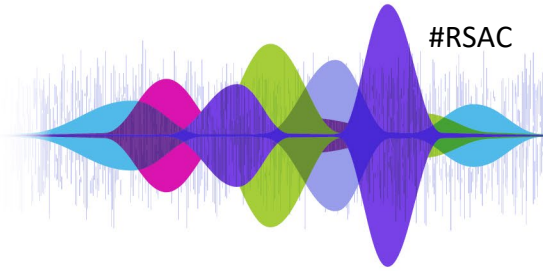
Proposed Changes to the Federal Rules of Evidence

Prof. Daniel Capra and Federal Rules Advisory Committee

Proposed New Fed. R. Evid. 707

Rule 707. Machine-generated Evidence

Where the output of a process or system would be subject to Rule 702 if testified to by a human witness, the court must find that the output satisfies the requirements of Rule 702 (a)-(d). This rule does not apply to the output of basic scientific instruments or routinely relied upon commercial software.



Proposed Changes to the Federal Rules of Evidence

Prof. Daniel Capra and Federal Rules Advisory Committee

Proposed New Fed. R. Evid. 901(c) for Disputed or Potentially Fabricated Evidence

901(c): Potentially Fabricated Evidence Created by Artificial Intelligence

If a party challenging the authenticity of computer-generated or other electronic evidence demonstrates to the court that a jury reasonably could find that the evidence has been fabricated, in whole or in part, by artificial intelligence, the evidence is admissible only if the proponent demonstrates to the court that it is more likely than not authentic.

RSAC 2025 Conference

State of technical solutions that judges and experts might use in court

Many Voices.
One Community.

Leveraging cross-industry partnerships

C2PA

Coalition for Content Provenance and Authenticity

Founded by Microsoft and Adobe, includes Arm, BBC, Intel, and Truepic.

Project Origin

Protecting Trusted Media

Content Authenticity Initiative

A community of media and tech companies, NGOs, academics, and others working to promote adoption of an open industry standard for content authenticity and provenance.

Establish a Foundation for Trust in Media

"Misinformation is a growing threat to the integrity of the information eco-system. Having a provable source of origin for media, and knowing that it has not been tampered with en-route, will help to maintain confidence in news from trusted providers."

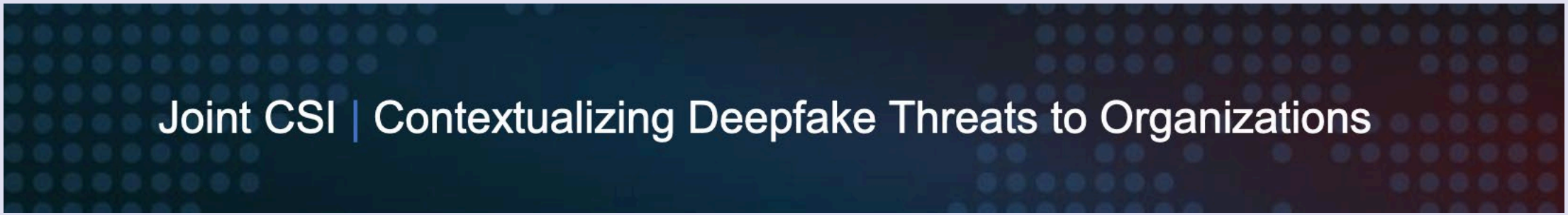
- **Deepfake Defense Tech Ready for Commercialization, Transition**

- https://www.darpa.mil/research/programs/semantic-forensics

Deepfake Detection Resources

- SANS Institute: **Learn a New Survival Skill: Spotting Deepfakes**
- MIT Media Lab: **Detect DeepFakes: How to counteract misinformation created by AI**
- Microsoft: **Spot the Deepfake**
- Northwestern U: **Deepfake-detection system**
  Free platform to help journalists determine validity of digital artifacts

FBI Recommendations for Resisting Deepfakes



(1) Select and implement technologies to detect deepfakes and demonstrate media provenance, including:

- Real-time verification capabilities
- Passive detection techniques
- Protection of high priority officers and their communications.

(2) Protect public data of high-priority individuals.

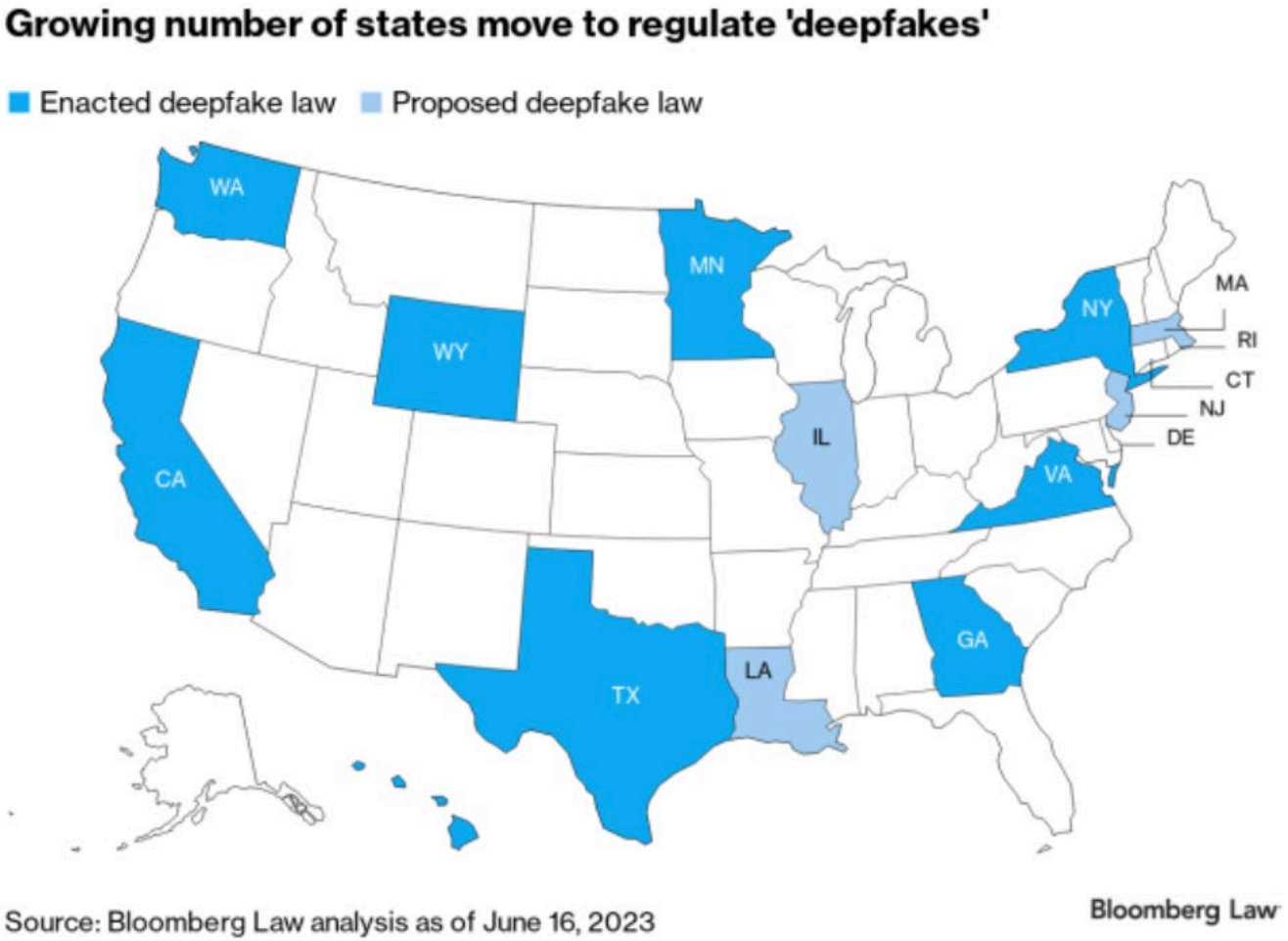
- Use active authentication techniques such as watermarks and/or CAI standards.
- Minimize the impact of deepfakes:
  - Plan and rehearse: Ensure plans are in place to respond to a variety of deepfake techniques
  - Train personnel

DEEPFAKE

PHASE TWO || MITIGATION MEASURES

Homeland Security

Nine states have enacted laws regulating deepfakes audio and visual forgeries created with AI



- California, Texas, Washington and Minnesota passed bills that prohibit using deepfakes to influence elections without clear disclosures; Minnesota's law also prohibits using deepfakes to make sexual images of a person without their consent.
- AB-972, CA 2022; SB-751, TX 2019; SB-5152, WA 2023; HF-1370, MN 2023
- New York bill (S1042A) makes it illegal to disseminate AI-generated explicit images or "deepfakes" of a person without their consent. Deepfake porn involves creating fake sexually explicit media using someone's likeness.

Topics/Learning Objectives

- **Understand the threat of deepfakes** and the looming threat to the courts as judges grapple with authenticity and reliability of AI-generated evidence, and the potential to erode public trust in the judicial system.
- Sort out the **serious threats and legitimate and creative uses of deepfakes**.
- Become familiar with the **legal and regulatory landscape**.
- Are current laws and regulations adequate to address evolving concerns about the malicious use of deepfakes? For the courts, do the Federal Rules of Evidence provide the tools judges need to address AI-generated deepfake evidence in a meaningful way?

Summary/Takeaways

- With a better understanding of the impact of deepfakes, take steps to protect yourself and your organization.
- Be vigilant in questioning the provenance of potentially deepfake information.
- Follow the NIST AI Risk Management Framework to evaluate deepfake and disinformation risks.
- Monitor laws and lawsuits to gain insights on risks.
- Adopt practical steps and use advanced security measures to protect against fraud.
- Stay updated on technology, as the tools to create and detect deepfakes are constantly evolving.

RSAC 2025 Conference

Resources

Resources

- Federal Rules of Evidence
- Fed. R. Evid. Advisory Committee

- *Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery*,
- Rebecca A. Delfino
- *The Deepfake Dilemma*, Hon. Paul Grimm

Resources

Paul W. Grimm, Maura R. Grossman, and Gordon V. Cormack, ***Artificial Intelligence as Evidence***, 19 Nw. J. Tech. & Intell. Prop. 9-106 (2021)

Cynthia Cwik, Paul W. Grimm, Maura R. Grossman, and Toby Walsh, ***Artificial Intelligence, Trustworthiness, and Litigation***, Artificial Intelligence and the Courts: Materials for Judges 1-35 (American Association for the Advancement of Science (2022))

Maura R. Grossman, Paul W. Grimm, Daniel G. Brown, and Molly (Yiming) Xu, ***The GPTJudge: Justice in a Generative AI World***, 23:1 Duke L. & Tech. Rev. 1-34 (Dec. 2023)

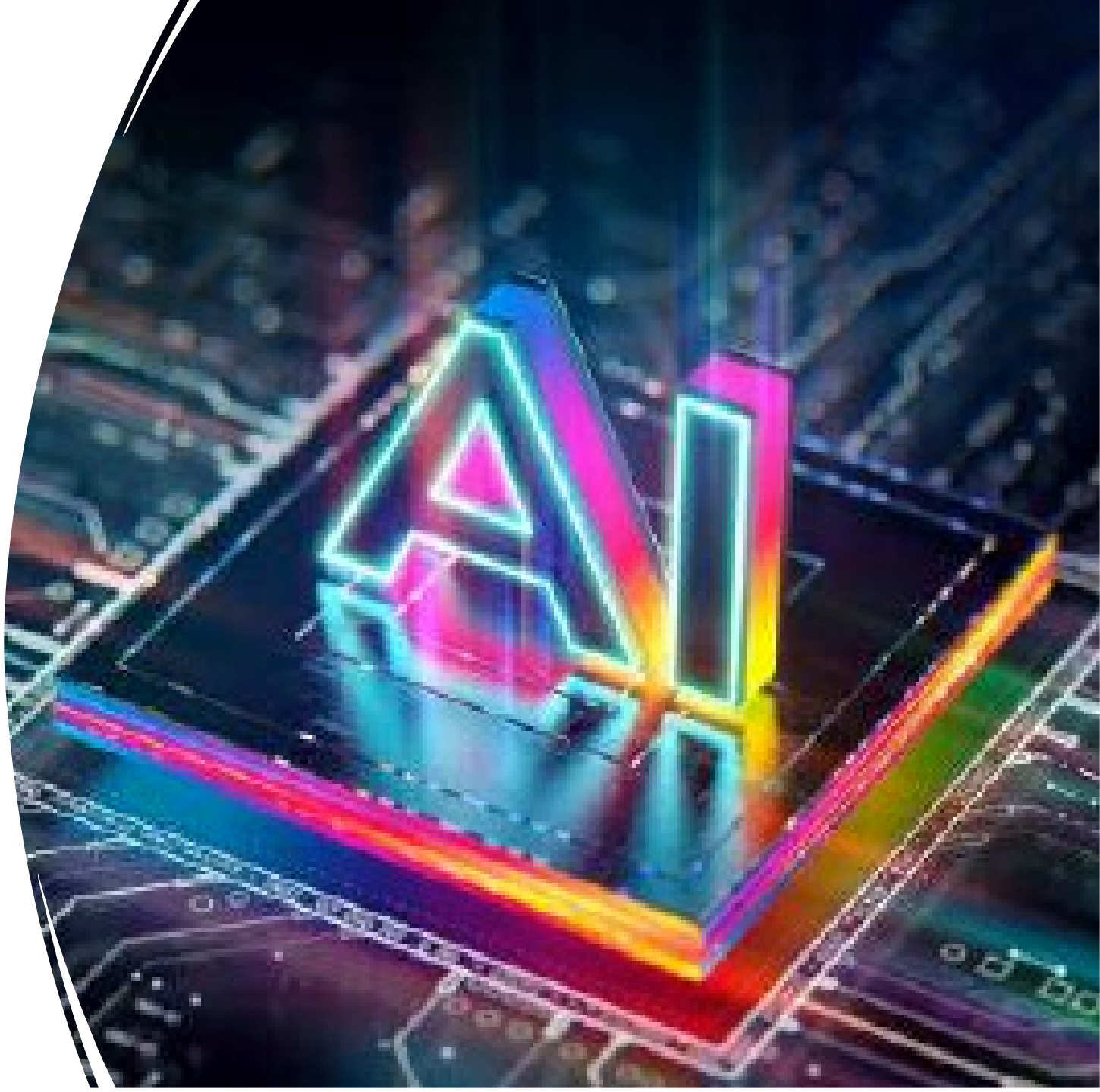
Abhishek Dalal, Chongyang Gao, Paul W. Grimm, Maura R. Grossman, Daniel W. Linna, Jr., Chiara Pulice, V.S. Subrahmanian, and John Turnheim, ***Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases***, The U. of Chicago Legal Forum 75-113 (2024)

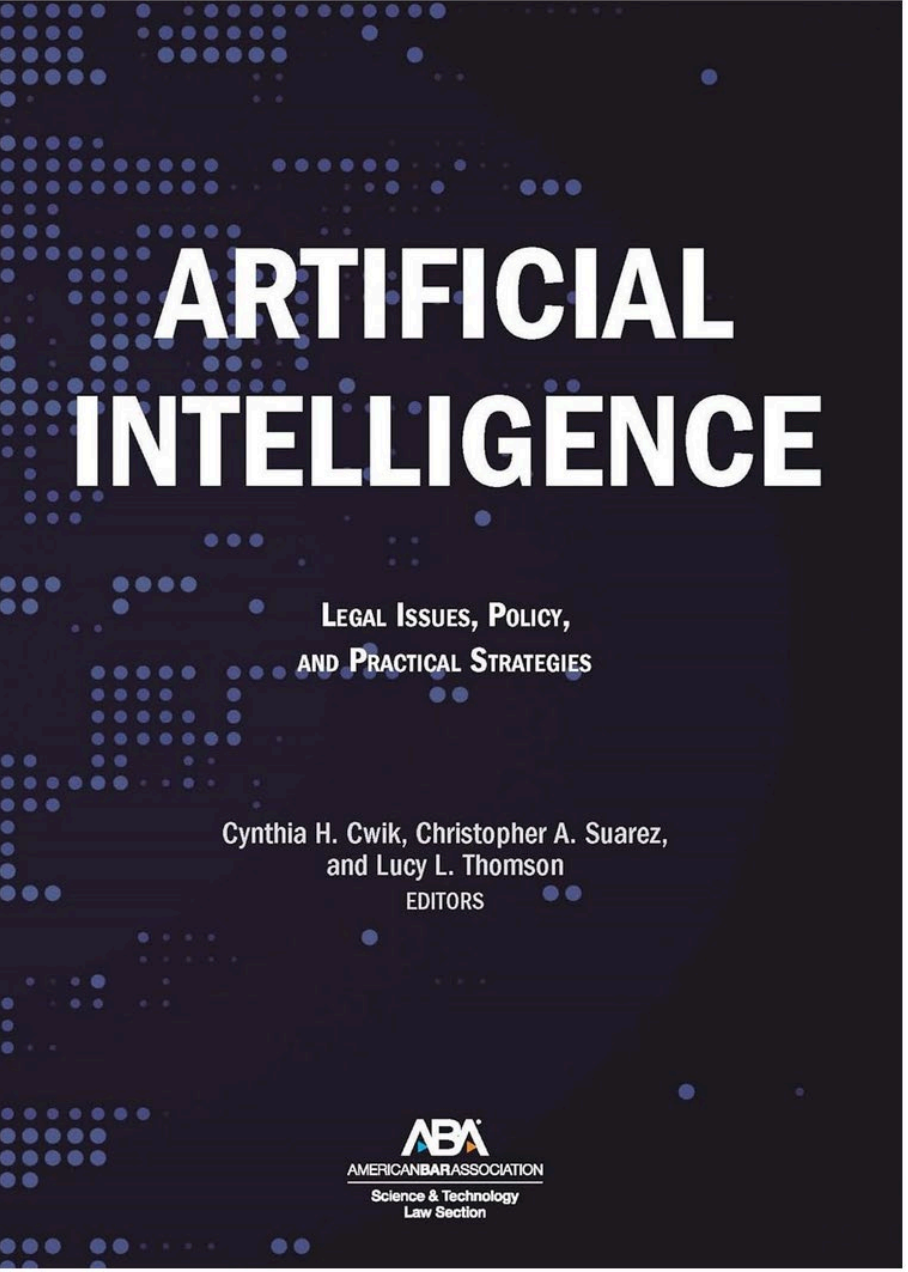
American Bar Association Task Force on Law and Artificial Intelligence

Year I Report on the Impact of AI on the Practice of Law (2024)

ambar.org/aiLaw

- Law Practice and Ethical Dilemmas
- AI and the Courts
- Access to Justice
- Legal Education
- Governance
- Trustworthy and Responsible AI/ Risk Management



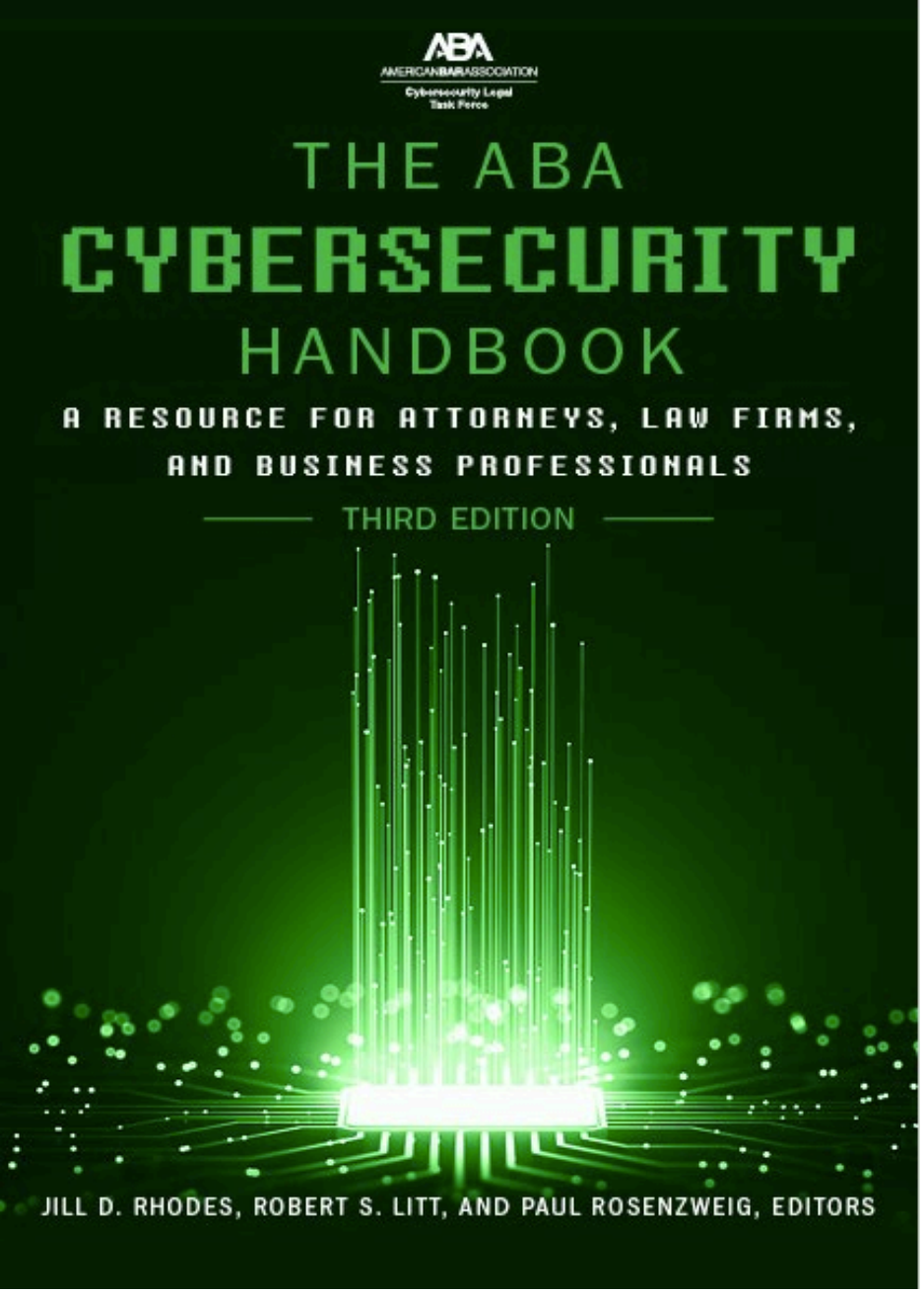


ABA Best-selling Book – in RSAC bookstore

AI is rapidly transforming our world and the practice of law. AI will improve many aspects of legal practice, but it also presents significant challenges. This book:

- Contains reflections and chapters by over 40 preeminent AI and legal experts from in-house legal departments, private practice, academia, government, and the judiciary.
- Covers a wide range of important topics concerning AI and the law, including the need to protect confidential client information, ensure accurate outputs, and mitigate privacy, security, and IP risks.
- Provides practical advice on how to navigate this complex and rapidly evolving technology.

Order at **ambar.org/aibook**.
Use code **SPTNCI20** **for a** **20% discount**



Recognizing the growing volume and sophistication of cyberattacks, *The ABA Cybersecurity Handbook* addresses the current overarching cyber threats, describes how the technology works, outlines key legal requirements and ethical issues, and highlights special considerations for lawyers and practitioners of all types.

From the ABA Cybersecurity Legal Task Force (ambar.org/cyber), this award-winning bestseller has over two dozen contributing authors.

Order at **ambar.org/cybersecurity**.
Use code **PETNI50 for a 50% discount**

RSAC™ 2025 Conference

Many Voices.
One Community.

Thank You